TECOLOTE RESEARCH, INC.

Bridging Engineering and Economics
Since 1973

Examination of Functional Correlation And Its Impacts On Risk Analysis

Alfred Smith

NASA Conference

June 2007

■ Los Angeles ■ Washington, D.C. ■ Boston ■ Chantilly ■ Huntsville ■ Dayton ■ Santa Barbara
■ Albuquerque ■ Colorado Springs ■ Columbus ■ Ft. Meade ■ Ft. Monmouth ■ Montgomery ■ Ogden
■ Patuxent River ■ Pensacola ■ San Diego
■ Charleston ■ Cleveland ■ Denver ■ New Orleans ■ Oklahoma City ■ Silver Spring ■ Tacoma ■
Warner Robins AFB ■ Vandenberg AFB

Overview

- Define “Applied Correlation”
- Using a simulation tool to replicate published results of applied correlation impact on throughputs
- A new twist on a well know chart: potential Std Dev underestimated if correlation left at zero
- Define “Functional Correlation”
- Comparing correlation applied to throughputs vs functionally correlated models
- Conclusions

Note: all simulations performed at 10,000 iterations, Latin Hypercube and all distributions truncated at 0

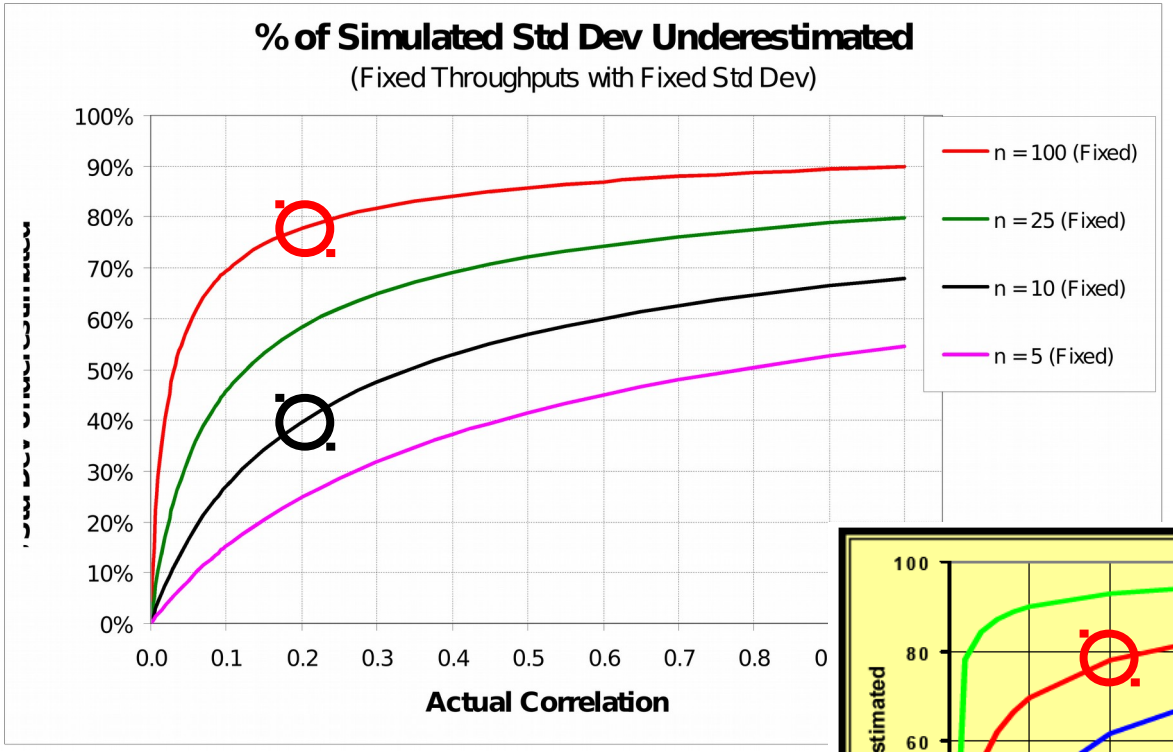
Defining Applied Correlation

- The correlation coefficient indicates the strength and direction (+ve or –ve) of a linear association between two random variables
- Simulation tools allow you to “apply” correlation between two or more uncertainty distributions
- Example illustrates 0.25 correlation “applied” to otherwise independent random variables
- Note that this model sums constant point estimates, with different distributions but with the same Std Dev

WBS Element	Point Estimate	Stdev/PE	Skew	Alpha	Beta
Item 1 LN	100.0 (50%)	0.25			
Item 2 Norm	100.0 (50%)	0.25			
Item 3 Tri	100.0 (50%)	0.25	Center		
Item 4 Beta	100.0 (50%)	0.25		0.5	0.5
Item 5 Unif	100.0 (50%)	0.25	Center		

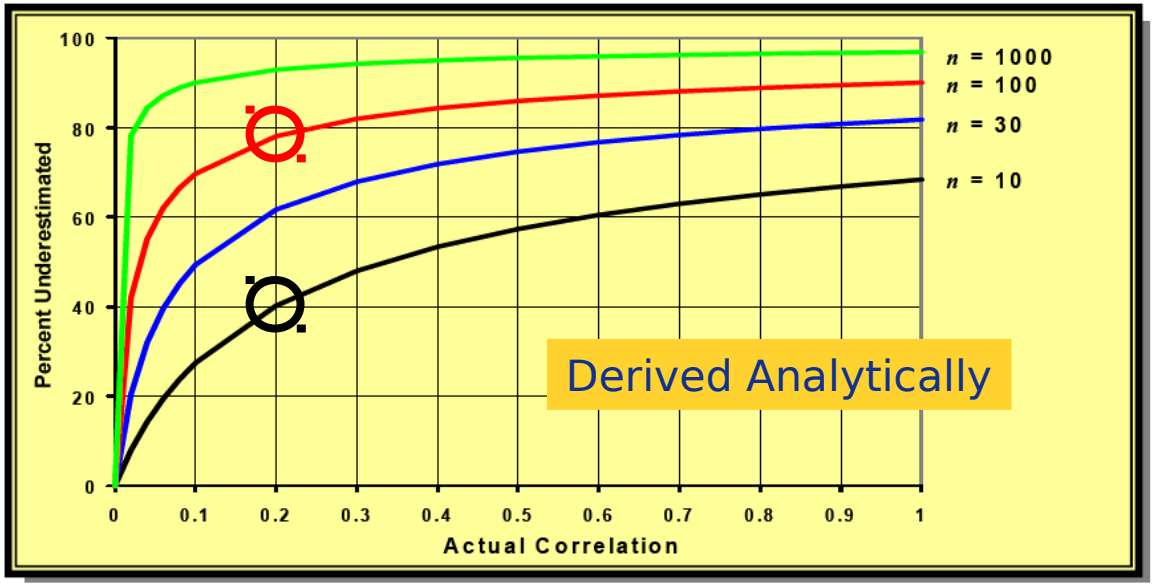
Item 1 LN	Item 2 Norm	Item 3 Tri	Item 4 Beta	Item 5 Unif
1.00	0.25	0.25	0.25	0.25
	1.00	0.25	0.25	0.25
		1.00	0.25	0.25
			1.00	0.25
				1.00

Well Known Correlation Impact on Sum of Throughputs



- **Based on: all throughputs of equal value and equal std dev**
- **Simulation employed five different distributions**
- **Simulation tool replicates published correlation impact**
- **Correlation importance increases with number of**

WBS/CES	Point Estimate	Mean	Std Dev
Item 1 LN	100.0 (50%)	102.9	25.0
Item 2 Norm	100.0 (50%)	100.0	25.0
Item 3 Tri	100.0 (50%)	100.0	25.0
Item 4 Beta	100.0 (50%)	100.0	25.0
Item 5 Unif	100.0 (50%)	100.0	25.0



From: *Why Correlation Matters in Cost Estimating*; Dr. Stephen A. Book; The Aerospace Corporation; 32nd DODCAS; 2-5 February 1999

Throughput Model With Various Point Estimates and Various Std Dev

	Point Estimate	Mode	Mean	Low	High or Std Dev	Alpha or Shape	Beta or Scale
Total 6 Elements	63,282.65		77,967.88				
Item 1 LN	**4,192.86**		**4,540.46**		**1,886.78**		
Item 2 Norm	**9,401.51**	9,401.51	**9,401.51**		**2,350.38**		
Item 3 Tri	**6,678.31**	**6,678.31**	8,681.80	**4,674.82**	**14,692.28**		
Item 4 Beta	**12,809.19**	12,809.19	13,724.13	**8,234.48**	**21,958.60**	**2**	**3**
Item 5 Unif	**19,624.29**	19,624.29	22,567.94	**7,849.72**	**37,286.15**		
Item 6 Weib	**10,576.50**	10,576.50	19,052.05			**1.529808**	**2**

- **Bold elements used to define distributions**
- **Non bold mode, mean calculated from standard equations**
- **Weibull Shape value selected to cause a point estimate of 1 to be the mode. This distribution is multiplied by the model point estimate.**

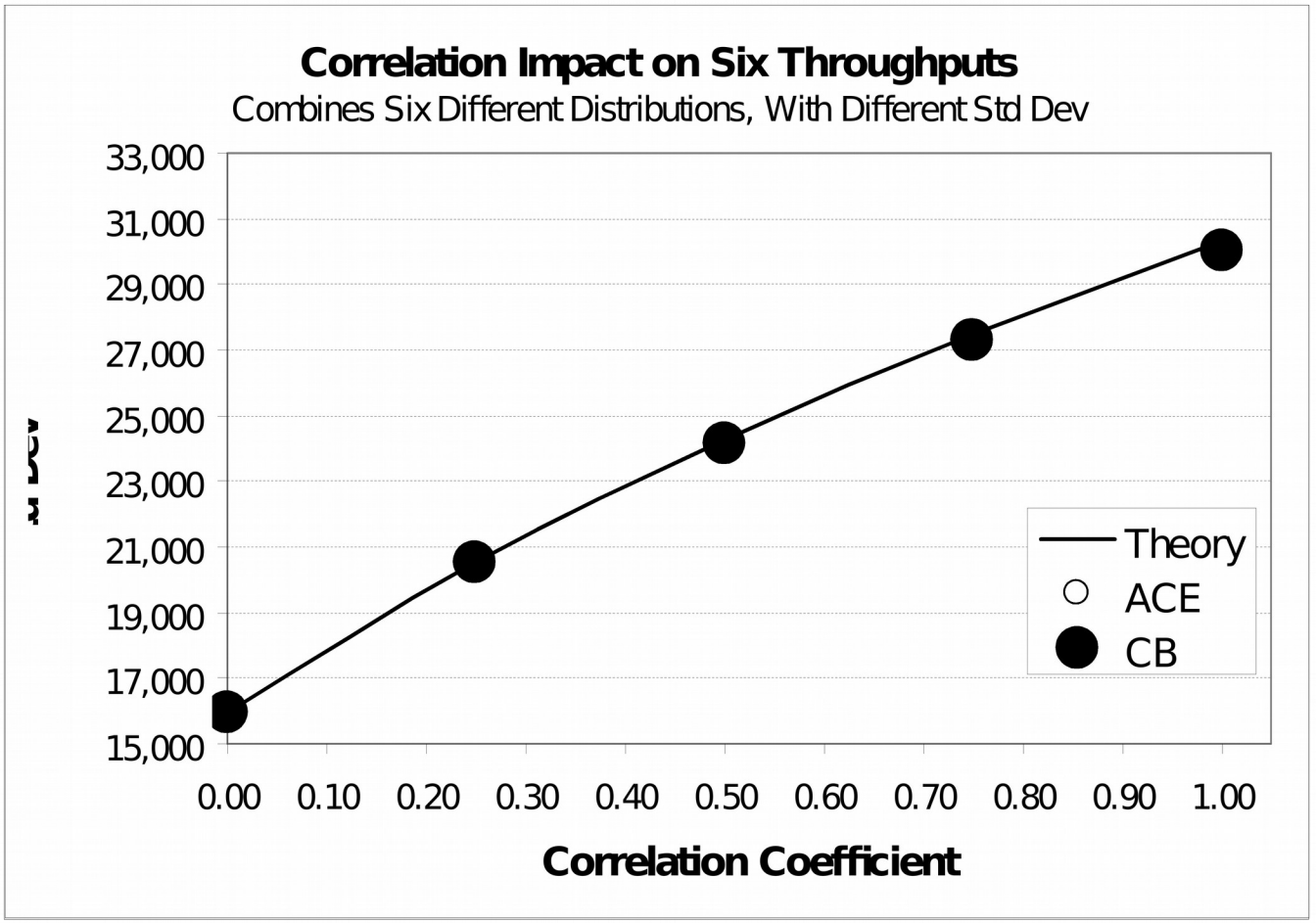
Variance Equations

Item 3 Tri	$\frac{(Max - Min)^2 + (Mode - Min)(Mode - Max)}{18}$
Item 4 Beta	$\frac{\alpha \times \beta}{(\alpha+\beta)^2 \times (\alpha+\beta+1)} \times (max - min)^2$
Item 5 Unif	$\frac{(max - min)^2}{12}$
Item 6 Weib	$b\left(\Gamma(1+\frac{2}{\alpha}) - \Gamma^2(1+\frac{1}{\alpha})\right)$

Corr = 0	CV	Theoretical StdDev	ACE Std Dev	CB Std Dev	ACE/ Theory	CB/ Theory
Total 6 Elements	0.20	15,965.28	15,957.30	15,956.05	100.0%	99.9%
Item 1 LN	0.42	1,886.78	1,888.39	1,887.51	100.1%	100.0%
Item 2 Norm	0.25	2,350.38	2,351.29	2,350.65	100.0%	100.0%
Item 3 Tri	0.25	2,164.02	2,164.34	2,164.12	100.0%	100.0%
Item 4 Beta	0.20	2,744.83	2,745.34	2,744.94	100.0%	100.0%
Item 5 Unif	0.38	8,497.57	8,498.00	8,498.00	100.0%	100.0%
Item 6 Weib	0.67	12,703.55	12,696.81	12,702.51	99.9%	100.0%

Using Simulation Tools to Study Impact of Correlation on Throughputs

$$\text{Total Cost Variance} = \sum_{k=1}^{n} \sigma_k^{\,2} + 2\sum_{k=2}^{n}\sum_{j=1}^{k-1} \rho_{jk}\sigma_j\sigma_k$$



Yes, Even 100 Elements Match Theory

Enter Override Here-->	0.5
Simulation/Theory =	98.65%

$$\text{Total Cost Variance} = \sum_{k=1}^{n} \sigma_k^{\,2} + 2\sum_{k=2}^{n}\sum_{j=1}^{k-1} \rho_{jk}\sigma_j\sigma_k$$

Model Std Dev	Theory Std Dev	Item 1 LN	Item 2 Norm	Item 3 Tri	Item 4 Beta	Item 98 Tri	Item 99 Beta	Item 100 Unif
16,564.45	16,791.68	85.26	125.65	167.47	312.2	149.04	415.44	127.02
85.26	Item 1 LN	**1.000**	0.500	0.500	0.500	0.500	0.500	0.500
125.65	Item 2 Norm	0.500	**1.000**	0.500	0.500	0.500	0.500	0.500
167.47	Item 3 Tri	0.500	0.500	**1.000**	0.500	0.500	0.500	0.500
312.20	Item 4 Beta	0.500	0.500	0.500	**1.000**	0.500	0.500	0.500
411.92	Item 5 Unif	0.500	0.500	0.500	0.500	0.500	0.500	0.500
420.50	Item 6 LN	0.500	0.500	0.500	0.500	0.500	0.500	0.500
117.95	Item 95 Unif	0.500	0.500	0.500	0.500	0.500	0.500	0.500
195.99	Item 96 LN	0.500	0.500	0.500	0.500	0.500	0.500	0.500
244.66	Item 97 Norm	0.500	0.500	0.500	0.500	0.500	0.500	0.500
149.04	Item 98 Tri	0.500	0.500	0.500	0.500	**1.000**	0.500	0.500
415.44	Item 99 Beta	0.500	0.500	0.500	0.500	0.500	**1.000**	0.500
127.02	Item 100 Unif	0.500	0.500	0.500	0.500	0.500	0.500	**1.000**

Calculation demonstrates even 100 element model with a variety of distributions (lognormal, triangular, normal, beta, uniform, weibull) returns a total std dev that matched theory

Using Final Simulated Correlations

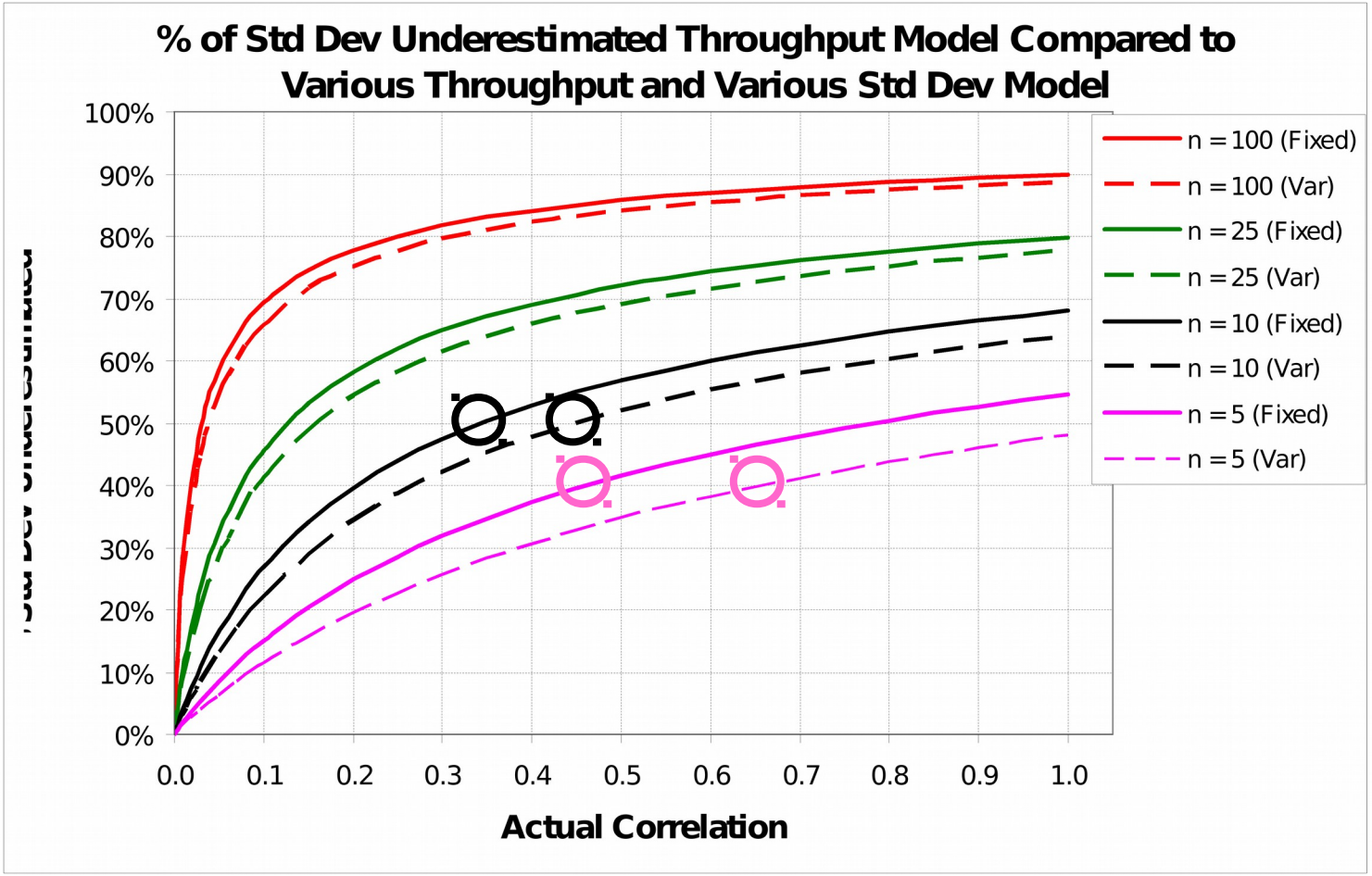
Enter Override Here-->	
Simulation/Theory =	100.00%

$$\text{Total Cost Variance} = \sum_{k=1}^{n} \sigma_k^{\,2} + 2\sum_{k=2}^{n}\sum_{j=1}^{k-1} \rho_{jk}\sigma_j\sigma_k$$

Model Std Dev	Theory Std Dev	Item 1 LN	Item 2 Norm	Item 3 Tri	Item 4 Beta	Item 98 Tri	Item 99 Beta	Item 100 Unif
16,564.45	16,564.43	85.26	125.65	167.47	312.2	149.04	415.44	127.02
85.26	Item 1 LN	**1.000**	0.486	0.486	0.489	0.483	0.474	0.460
125.65	Item 2 Norm	0.486	**1.000**	0.499	0.505	0.494	0.502	0.489
167.47	Item 3 Tri	0.486	0.499	**1.000**	0.492	0.497	0.500	0.478
312.20	Item 4 Beta	0.489	0.505	0.492	**1.000**	0.497	0.498	0.482
411.92	Item 5 Unif	0.467	0.490	0.492	0.477	0.484	0.489	0.480
420.50	Item 6 LN	0.487	0.481	0.481	0.472	0.479	0.482	0.468
117.95	Item 95 Unif	0.474	0.485	0.480	0.476	0.484	0.479	0.483
195.99	Item 96 LN	0.482	0.489	0.484	0.488	0.488	0.479	0.469
244.66	Item 97 Norm	0.484	0.501	0.501	0.504	0.492	0.493	0.494
149.04	Item 98 Tri	0.483	0.494	0.497	0.497	**1.000**	0.489	0.496
415.44	Item 99 Beta	0.474	0.502	0.500	0.498	0.489	**1.000**	0.476
127.02	Item 100 Unif	0.460	0.489	0.478	0.482	0.496	0.476	**1.000**

If you capture the simulation iterations and measure the Pearson Product correlation actually manifested by the simulation and use that correlation matrix, the std dev returned by the tool exactly matches theory

Impact if Throughputs and Std Dev are not Fixed



- **Suggests that defaults should be higher if you wish to protect against 50% underestimated**
- **For 10 elements, if you wanted to protect against 50% underestimated, you need to apply 0.45, not 0.35 (for 5 elements 40%, 0.65, not 0.45)**

Defining Functional Correlation

- **Functional correlation is correlation induced into a model through the algebra of the model**
- **Examples:**
  - Item 2 and 3 are functionally correlated through a common wgt variable
  - Item 2 and item 4 are functionally correlated through a factor relationship
  - Item 4 and 5 are functionally correlated through a common

WBS	PE	CV	Eq / Thruput	Form	Low	High
Total	1,482.2 (24%)	0.27				
Item 1	400.0 (36%)	0.40	400	Triangular	70%	180%
Item 2	338.6 (36%)	0.29	256.2+0.05682*Wgt^1.374	LogNormal		130%
Item 3	239.9 (35%)	0.47	30.15+1.049*Wgt	Normal		140%
Item 4	203.2 (43%)	0.62	.6*Item2	Normal		165%
Item 5	300.4 (43%)	0.66	3.5*(Item2+Item3)^0.7	LogNormal		180%
Weight	200.0 (24%)	0.31	200	Triangular	90%	180%

A Functionally Correlated Model

WBS	PE	CV	Eq / Thruput	Form	Low	High
Total	1,482.2 (35%)	0.24				
Item 1	400.00 (36%)	0.40	400	Triangular	70%	180%
Item 2	338.64 (50%)	0.26	256.2+0.05682*Wgt^1.374	LogNormal		130%
Item 3	239.95 (50%)	0.38	30.15+1.049*Wgt	Normal		140%
Item 4	203.18 (49%)	0.60	.6*Item2	Normal		165%
Item 5	300.42 (50%)	0.64	3.5*(Item2+Item3)^0.7	LogNormal		180%
Weight	200		200			

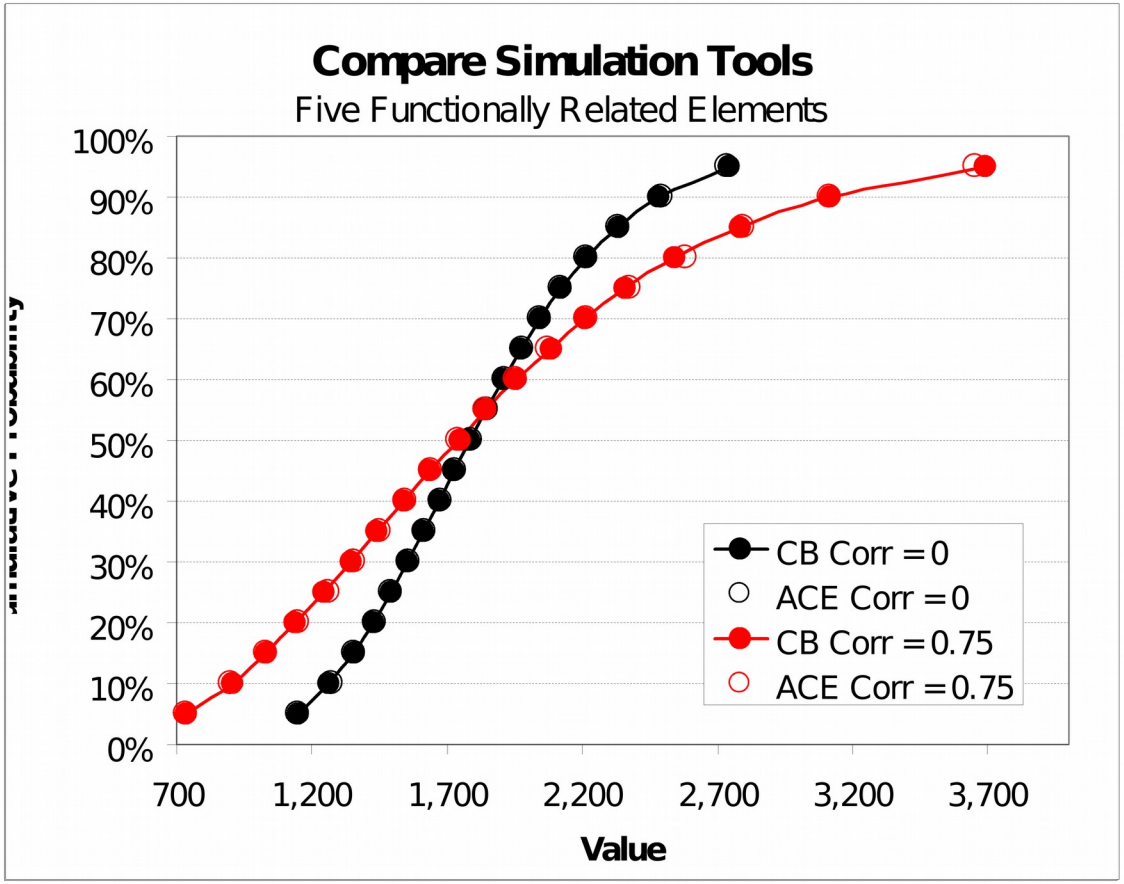
No Uncertainty on Weight					
Item	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	1.00	0.00	0.00	0.00	0.00
Item 2		1.00	-0.01	**0.44**	**0.18**
Item 3			1.00	0.00	**0.16**
Item 4				1.00	0.10
Item 5					1.00

WBS	PE	CV	Eq / Thruput	Form	Low	High
Total	1,482.2 (24%)	0.27				
Item 1	400.0 (36%)	0.40	400	Triangular	70%	180%
Item 2	338.6 (36%)	0.29	256.2+0.05682*Wgt^1.374	LogNormal		130%
Item 3	239.9 (35%)	0.47	30.15+1.049*Wgt	Normal		140%
Item 4	203.2 (43%)	0.62	.6*Item2	Normal		165%
Item 5	300.4 (43%)	0.66	3.5*(Item2+Item3)^0.7	LogNormal		180%
Weight	200.0 (24%)	0.31	200	Triangular	90%	180%

Uncertainty on Weight					
Item	Item 1	Item 2	Item 3	Item 4	Item 5
Item 1	1.00	0.00	0.00	0.00	0.00
Item 2		1.00	**0.26**	**0.48**	**0.24**
Item 3			1.00	**0.13**	**0.25**
Item 4				1.00	0.14
Item 5					1.00

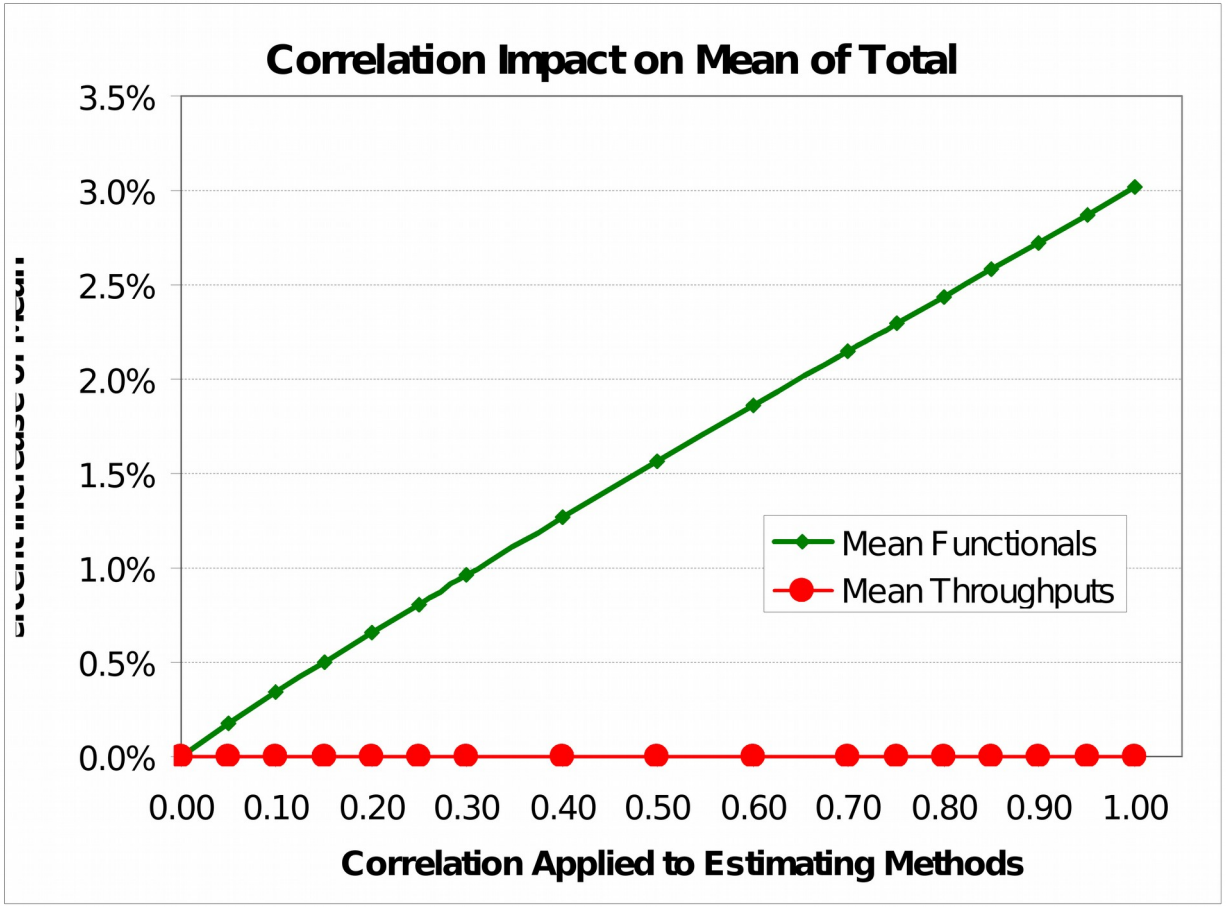
- **Item 2 and 3, and Item 3 and 4 are not correlated when weight is "certain"**
- **Item 2 and 3 and Item 3 and 4 become "correlated" when weight variable (common to item 2 and 3) is made uncertain**
- **Note that CVs increase and item 2 and 5, 3 and 5 correlation increases**

Simulation Tools Capture Functional Correlation



- **Must ensure CERs are driven from forecasts!**
- **Applying correlation does have an impact on already functionally correlated items**

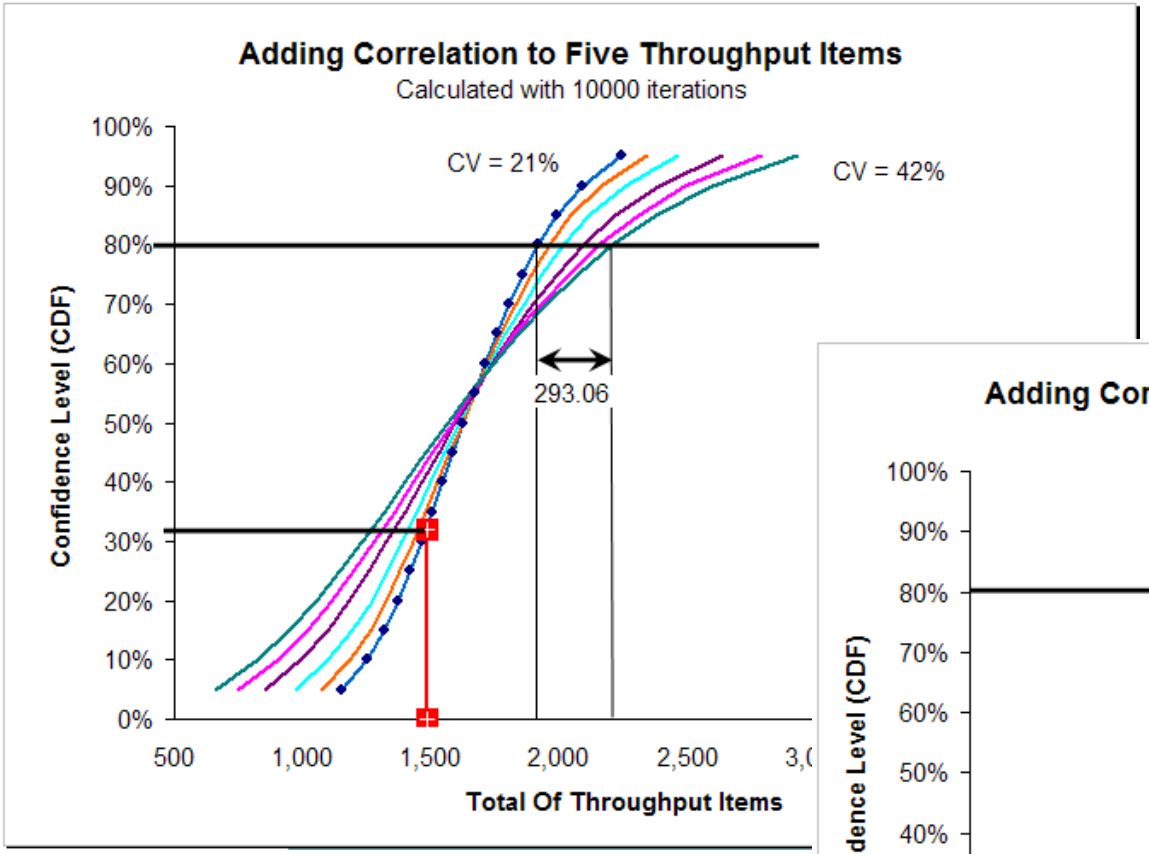
Functional Correlation Affects the Mean!



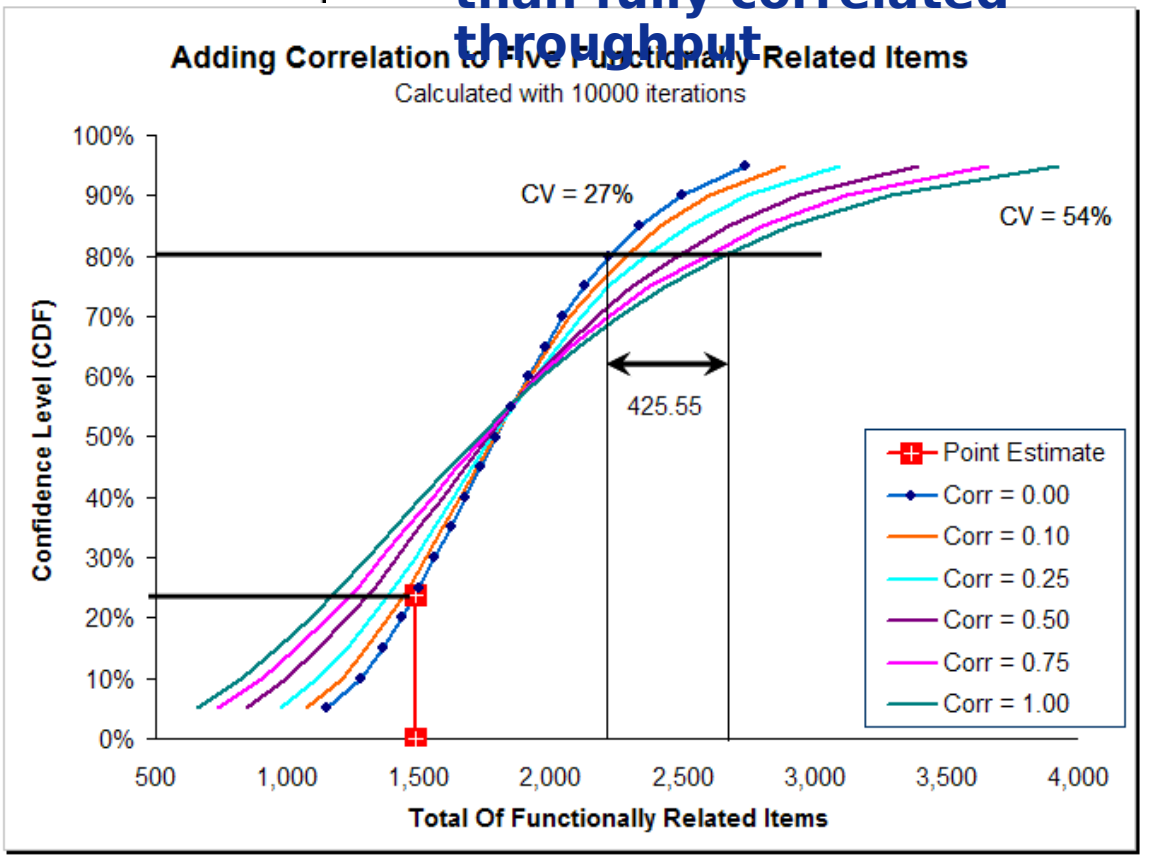
- **In this model, the mean of the estimate increases linearly with correlation (albeit by only a few percent)**
- **The mean of throughputs is NOT affected by correlation**

Pointed out to the author by Erik Burgess as a result of a review of the AFCAA Cost Risk Handbook

Correlation Impact On Throughputs vs Functional

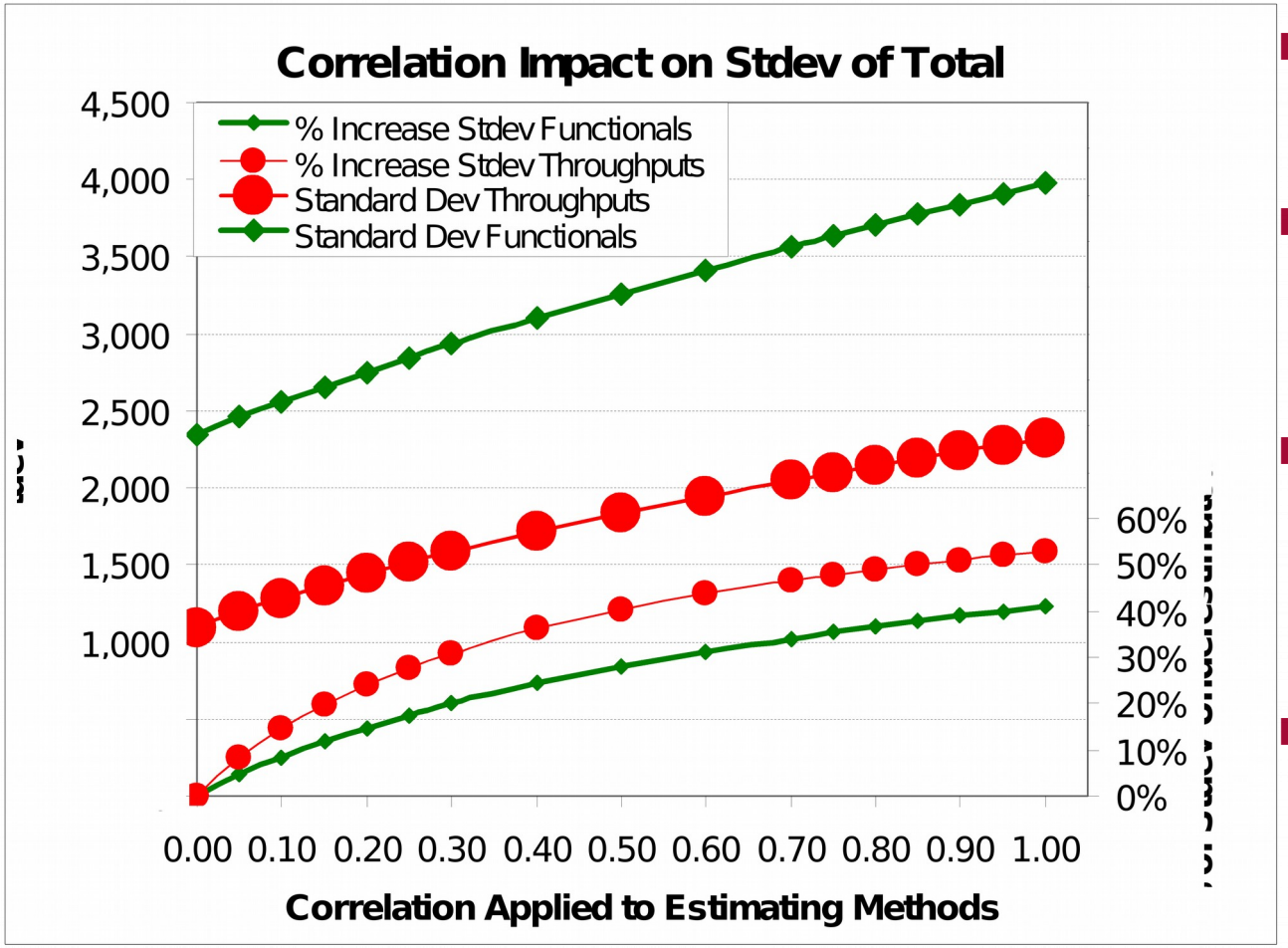


- **80% from functional is greater than fully correlated throughput**
- **Uncorrelated functional has greater variance than fully correlated throughput**



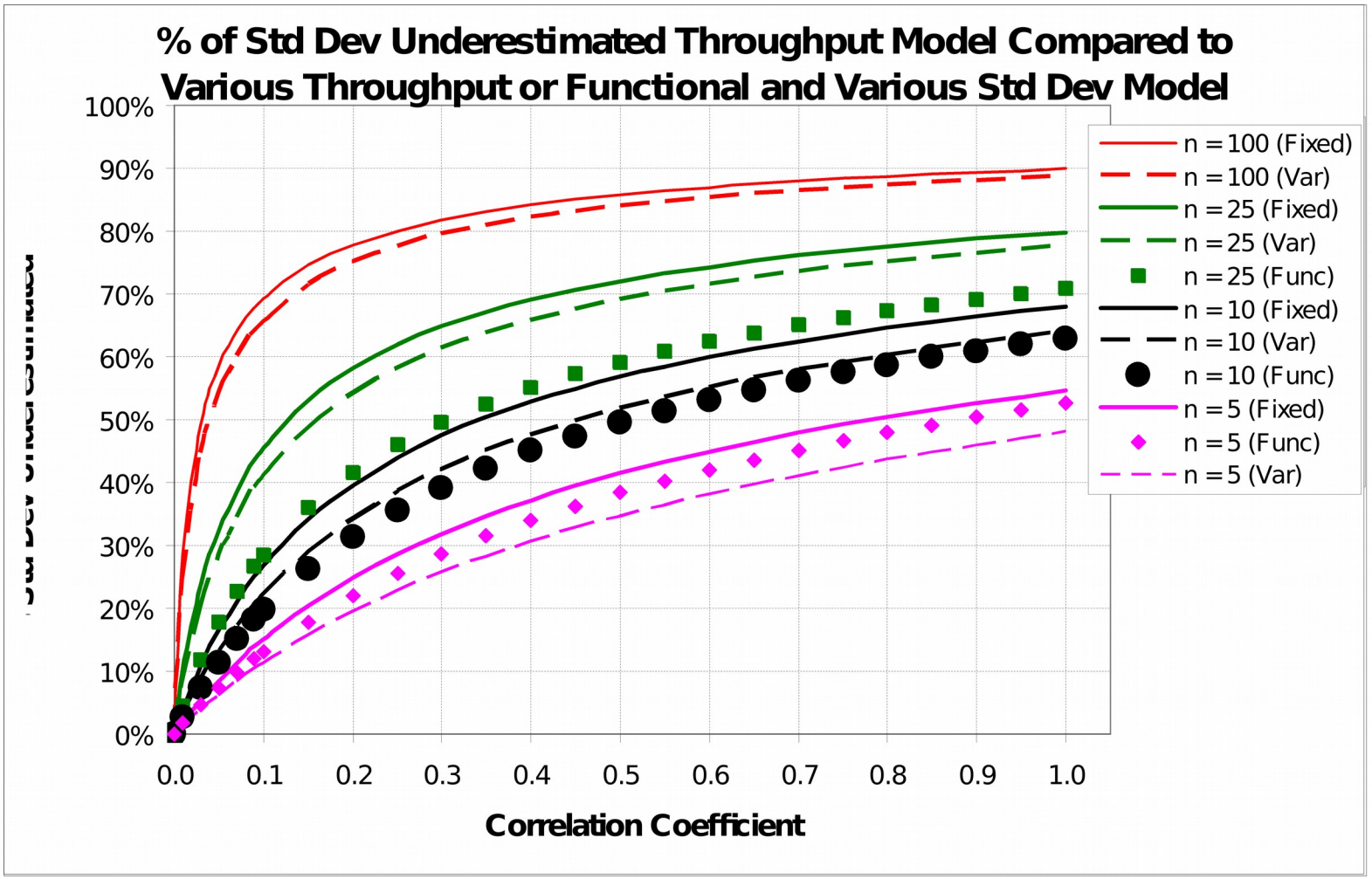
- **PE from throughput model between 32-45%**
- **PE from functional model between 23-40%**

Correlation Impact On Throughputs vs Functional



- **Identical point estimate for throughputs and functional version**
- **Uncorrelated functional has greater variance than fully correlated throughput**
- **Potential for underestimating is LESS (in relative terms) if model is functionally correlated**
- **Potential for underestimating is MORE (in absolute terms) if model is functionally correlated**

Impact if Items Functionally Related and Std Devs are not Fixed



In and Out of Trouble

Heading for Trouble

- **Throughput (number) PE**
  - May miss important relationship that functional correlation would normally capture
  - Simulate by applying correlation
- **Ignoring correlation**
  - Uncertainty distributions aren't enough
  - Variance at total will be underestimated
- **Layering matrix atop Funtional**
  - Correlation may already exist due to functional relationship
  - Assigning additional input coefficient will exaggerate impact of inputs
- **Reusing input driver**
  - Produces undesired "incidental" correlation due to common inputs
  - Increases variance at total

Escaping Trouble

- **Generate resulting correlations**
  - Run the model after defining distributions to find existing functional correlation
- **Study relationships**
  - Watch for unexplained FC – a symptom of shared drivers
  - Watch for low correlation among similar elements
- **Adjust input matrix**
  - Increase 0.0 to 0.25
  - Increase correlations among technically related throughputs
  - Adjust correlations between cost methods were there is evidence existing correlation is insufficient
- **Repeat**

Conclusions

- Simulation tools adequately capture the impact of correlation on both throughputs and functionally correlated models
- Functional correlation is correlation induced into a model through the algebra of the model
- Functional correlation affects the mean at the total level, correlation on throughputs does not
- Functional relationships can introduce unintended correlation (i.e. the same uncertain variable used across many cost methods)
- Functional correlation alone may establish a variance (with no applied correlation) that even fully correlated throughputs cannot achieve
- For 2 to 25 elements, defaults to capture underestimated variance when your model has varying throughputs and varying std dev (i.e. all the time) should be greater than previously published
- Build in functional relationships where ever you can!

TECOLOTE RESEARCH, INC.

References

- [1] Why Correlation Matters in Cost Estimating; Dr. Stephen A. Book; The Aerospace Corporation; 32nd DODCAS; 2-5 February 1999
- [2] Simulating Correlated Random Variables; Philip M. Lurie and Matthew S. Goldberg; Institute for Defense Analyses; 32nd DODCAS; 2-5 February 1999
- [3] 32. Impact Of Correlating CER Risk Distributions On A “Realistic” Cost Model; A Smith, Dr. Shu-Ping Hu; Tecolote Research; ISPA/SCEA Conference (Orlando) June 2003
- [4] Correlations in Cost Risk Analysis, Covert, R., SCEA Conference, Tysons Corner, VA, June 13-16, 2006
- [5] Cautionary Notes on Defining Parent-Level Correlations, Hu, Shu-Ping, 2006, White Paper, Tecolote Research